

April 4, 2025

Sent via ECF

Hon. Ona T. Wang
Daniel Patrick Moynihan
United States Courthouse
500 Pearl Street
New York, NY 10007-1312

RE: *The New York Times Co. v. Microsoft Corp., et al.* No. 1:23-cv-11195
New York Daily News, LP, et al. v. Microsoft Corp., et al., No. 1:24-cv-3285
The Center for Investigative Reporting, Inc. v. OpenAI, Inc., No. 1:24-cv-04872

Dear Judge Wang:

On behalf of all parties in the above-referenced actions, counsel for The New York Times Company hereby submits the below updated chart containing a summary of discovery disputes for discussion at the upcoming April 10, 2025 conference.

Given the number of issues included in the chart below, the parties have endeavored to list the issue categories in order of priority, alternating between Plaintiffs and Defendants.

Sincerely,

/s/ Ian Crosby
Ian Crosby

Key: Green represents Plaintiffs' filing; Purple represents Defendants' filing

Motion Dkt. No.	Date Filed	Opp. Dkt. No.	Topic in Dispute	Party Raising Dispute	Party Responding	Status of Dispute	Parties' Shorthand Positions
I. Disputes Related to Depositions							
382 Dkt. 430	1/13/25	403	Deposition Coordination Protocol	Defendants	News Plaintiffs	<p>In advance of the October 30 discovery conference, the parties submitted letter briefs about their competing proposals for a deposition protocol. <i>See</i> Dkts. 261, 278. At the October 30 conference, this Court stated this dispute was premature. Oct. 30, 2024 Tr. at 8:21-23.</p> <p>Prior to the January 22, 2025 conference, Defendants submitted a letter brief regarding a revised proposal and Plaintiffs responded</p>	<p>Defendants' Position: On Thursday, April 3, the Judicial Panel on Multidistrict Litigation (JPML) issued an order centralizing twelve actions, including these cases and other cases pending in this District and the Northern District of California, before Judge Stein. JPML Case No. 1:24-cv-03285, MDL No. 3143, Dkt. 42. As the JPML noted, "[t]he parties have been unable to agree on a deposition coordination protocol to eliminate duplicative depositions of defense witnesses." <i>Id.</i> at 3. Defendants look forward to working with the Court and plaintiffs in all of these centralized actions to "eliminate the potential for duplicative discovery" and ensure that Defendants' witnesses are deposed one time in an efficient manner.</p> <p>Given the JMPL's Order, there is no longer any basis for a deposition protocol that includes only a subset of the actions that have been centralized in this Court. Going forward, any deposition protocol should include all actions that have been centralized in these MDL proceedings, and as such the Defendants do not believe this is ripe for discussion at the upcoming status conference.</p> <p>Plaintiffs' position: The News Plaintiffs respectfully ask the Court to set a deadline for the parties to submit final proposals for a deposition protocol. The parties (including class plaintiffs) are close to resolving all of their disputes, as reflected in the term sheet the parties jointly prepared during the March 7 settlement conference. To the extent Defendants contend that material disagreements remain, the parties should continue discussing those issues and, if necessary, submit competing final proposals, for resolution by the Court by a date certain in April.</p>

Motion Dkt. No.	Date Filed	Opp. Dkt. No.	Topic in Dispute	Party Raising Dispute	Party Responding	Status of Dispute	Parties' Shorthand Positions
						with a counter-proposal. Dkts. 382, 403. The Court thereafter supervised two settlement conferences, on January 31, and on March 7.	
Dkt. 500	March 31, 2025	506	OpenAI's designation of Greg Brockman as an "apex" witness and limitations on deposition time	News Plaintiffs	OpenAI	The dispute is fully briefed and the Court has not yet heard argument.	<p>Plaintiffs' Position: News Plaintiffs seek an order confirming that OpenAI President Greg Brockman is not subject to the "apex" doctrine and therefore must sit for a standard deposition in this case. Here, that means Mr. Brockman should sit for 12 hours across all of (i) the three News Cases, (ii) the <i>Authors Guild</i> SDNY class action, and (iii) the N.D. Cal. Class action. (The parties have agreed to a default 12-hour time limit for any OpenAI employee who is deposed in all of these cases, and that time limit should apply to Mr. Brockman.)</p> <p>Documents produced by OpenAI prove that Mr. Brockman is "a far cry from the prototypical apex witness, who sits removed from the key facts of the dispute." <i>Oakley v. MSG Networks, Inc.</i>, 2024 WL 4134903, at *3 (S.D.N.Y. Sept. 10, 2024). For example, Mr. Brockman personally led OpenAI's efforts to obtain NYT content and to study how OpenAI's products compete with news companies. Dkt. 500-6; Dkt. 500-12. At deposition, another OpenAI employee confirmed Mr. Brockman's active involvement in relevant issues. Dkt. 500-5 (testimony from OpenAI employee Shantanu Jain).</p> <p>This dispute is ripe because OpenAI has said it will only present Mr. Brockman for eight hours across all of the cases, and that his deposition must take place in April, before the close of fact discovery in N.D. Cal.</p>

Motion Dkt. No.	Date Filed	Opp. Dkt. No.	Topic in Dispute	Party Raising Dispute	Party Responding	Status of Dispute	Parties' Shorthand Positions
							<p>OpenAI's Position: The News and S.D.N.Y. Class Plaintiffs ask this Court to order Greg Brockman, co-founder and current President of OpenAI, to sit for twelve hours of deposition over two days. But they have failed to justify such an intrusion on the time of Mr. Brockman, an Apex witness whose deposition is “disfavored” and given “special scrutiny” in this circuit. As President of OpenAI, Mr. Brockman leads critical aspects of OpenAI’s research, technical, and product efforts. Diverting him away from these Apex responsibilities for multiple days of deposition would be highly disruptive to OpenAI’s business. OpenAI’s proposal for Mr. Brockman to sit for an eight-hour deposition in one day for the News, S.D.N.Y. Class, and N.D. Cal. Class cases is beyond that required under the Apex doctrine and would allow each plaintiff group to explore any issues unique to their cases without improper and inefficient duplication precluded by Rule 26.</p> <p>Plaintiffs offer no compelling justification for their proposal. Their motions demonstrate significant overlap in the issues that they want to explore with Mr. Brockman—such as his role at OpenAI and his alleged involvement in developing GPT services and identifying datasets—that do not justify having him sit for twelve hours across two days. Moreover, many of the issues Plaintiffs cite can be explored with other OpenAI fact witnesses or 30(b)(6) designees—Mr. Brockman has not and will not be designated—including OpenAI’s relationship with Microsoft and whether OpenAI’s products compete with products from news companies like The Times. Finally, Plaintiffs materially misstate the confidential discussions among the parties in connection with two settlement conferences facilitated by this Court. No agreement on deposition coordination was reached, and OpenAI certainly never agreed to present Mr. Brockman or any Apex witness for 12 hours each.</p>
II. Disputes Related to Custodial Documents							
497 - Public 499 - Sealed	3/31/25	509	OpenAI’s request that The Times add three ESI custodians	OpenAI	New York Times	OpenAI had previously submitted letter briefs regarding The Times’s ESI custodians in November 2024 and	<p>OpenAI’s position: OpenAI seeks an order compelling the Times to add David Rubin, Michael Greenspon, and Sam Dolnick as custodians (renewing its previous motions, ECF 316, 393). At two prior discovery conferences, the Times has represented to OpenAI and the Court that its forthcoming productions from other custodians would render productions from Messrs. Rubin, Greenspon, and Dolnick duplicative.</p>

Motion Dkt. No.	Date Filed	Opp. Dkt. No.	Topic in Dispute	Party Raising Dispute	Party Responding	Status of Dispute	Parties' Shorthand Positions
						January 2025. ECF 316, 330 (briefing for December discovery conference); ECF 393, 408 (briefing for January discovery conference). The Court denied as premature both motions (ECF 365, 441).	<p>Time has revealed these assurances to be false. A review of the Times's productions to date shows that the overlap between its existing custodians and the three proposed by OpenAI is small. The Times has represented that its production is near complete. However, the 39,329 documents produced so far for the existing custodians that the Times claims cover the same ground has captured only 945 emails that involve the proposed custodians in any way—less than 2.5% of their production. It is clear from that figure alone that the proposed custodians possess nonduplicative documents. The Times argues that the “limited” number of documents produced from the proposed custodians indicates their documents are nonresponsive. But tellingly, The Times has refused to even provide search term hit counts for these custodians in support of that point. Moreover, the Times's productions show that each of these three individuals possesses uniquely relevant documents that have not been produced. For example, Rubin worked on presentations to upper management that brand risk stemming from GenAI. See ECF 499 at 2 (citing NYT_00706693). No current custodian properly captures the relevant scope of traditional licensing work performed by Greenspon. The Times reads ECF 352 much too broadly; it does not relieve them of their burden of producing documents regarding their allegations of market harm in traditional licensing markets. See id. at 2. And Dolnick oversaw GenerativeAI across the company across a period of time that neither replacement custodians Hardiman or Seward adequately cover. Id. at 3.</p> <p>Plaintiff's position: The three custodians OpenAI requests—on top of the 22 custodians already designated in this case—are no more appropriate now than they were in January. OpenAI repeatedly points to the limited number of documents from these specific individuals in The Times's existing productions as somehow proving that they have “uniquely relevant documents,” rather than accepting the far more plausible explanation that these individuals do not appear on more of the documents from The Times's custodians because—as The Times has repeatedly explained—they are not the individuals most likely to be part of responsive exchanges.</p> <p>More specifically, David Rubin's high-level role in overseeing brand strategy, subscriptions, advertising, licensing, books, film, television, commerce, and live events is too far attenuated to yield relevant discovery on the specific brand and reputation issues in this case. And</p>

Motion Dkt. No.	Date Filed	Opp. Dkt. No.	Topic in Dispute	Party Raising Dispute	Party Responding	Status of Dispute	Parties' Shorthand Positions
							<p>OpenAI's claim that the fact that a presentation on which Mr. Rubin comments has not yet been produced shows "it was in neither Frumin nor Han's files" is incorrect. Dkt. 499 at 2. The cited presentation was collected but was undergoing privilege review, and is now queued for production. With respect to Michael Greenspon, The Times has already designated four custodians with knowledge about licensing negotiations and the licensing market. It has also agreed to produce documents sufficient to show the key terms of relevant licenses (Gen AI and non-AI) from January 1, 2018, to the present. And, the Court has ruled the remainder of OpenAI's request—<i>viz.</i>, communications regarding licensing the asserted works for purposes other than LLM training—to be irrelevant. <i>See</i> Dkt. 352. Finally, with respect to Sam Dolnick, The Times has already designated five custodians specifically related to Gen AI: Zach Seward (the head of the AI Initiatives program), Sam Felix and Rebecca Grossman Cohen (the leads for Gen AI licensing), Alex Hardiman (the lead for AI products), and Barrett Sheridan (the Vice President for Strategy and Operations). OpenAI has not demonstrated why these custodians are insufficient.</p> <p>Finally, OpenAI's assertion that The Times has refused to "provide search term hit counts for these [proposed] custodians" is a sideshow. To provide such hit counts, The Times would have to engage in a full device and data collection for each individual. Such a costly request is clearly unwarranted here.</p>
494 – Public 495 - Sealed	3/31/25	510	Date range for Google Workspaces and Slack for six Times custodians	OpenAI	New York Times	OpenAI raised the issue of The Times's narrow custodial date ranges in connection with the October 30, 2024 and January 22, 2025 discovery conferences. The Court ordered the	<p>OpenAI's position: OpenAI seeks an order compelling The Times to expand the date range for its Google Workspace and Slack collection for six key Times custodians back to 2018, when The Times alleges the first acts of infringement occurred. These six custodians are likely to have pre-2022 documents relevant to at least two important issues. First, the custodians are likely to have relevant information regarding when The Times became aware that OpenAI had used a filtered version of Common Crawl (which included Times articles) to train its GPT-3 model, which OpenAI disclosed no later than July 22, 2020. This information is critical to OpenAI's statute of limitations defense, which Judge Stein recently ruled was subject to disputed questions of fact. Second, the custodians are likely to have relevant information regarding the alleged impact of LLMs on The Times's business model. The Times cannot claim that documents about this topic do not exist in this time period if it has not collected or performed a full search for them, and its own reporting shows that The Times was evaluating GPT-3's writing capabilities and the potential</p>

Motion Dkt. No.	Date Filed	Opp. Dkt. No.	Topic in Dispute	Party Raising Dispute	Party Responding	Status of Dispute	Parties' Shorthand Positions
						<p>parties to further meet-and-confer. OpenAI diligently pursued meet-and-confer, but The Times has not changed its position. OpenAI and The Times have now filed additional letter briefs on which the Court has not yet heard argument.</p>	<p>implications for The Times's journalists by at least July 2020. The Times has failed to substantiate its claims that this limited additional collection would be unduly burdensome. Indeed, those claims run headlong into the testimony of its custodial 30(b)(6) designee, who testified in deposition that The Times uses an industry standard collection tool—Google Vault—to collect these materials from Google Workspaces (email and Google Drive) used by Times employees. Contrary to The Times' assertions, Vault can be used to perform more complex, boolean searches. See https://support.google.com/vault/answer/2474474.</p> <p>Plaintiff's position: As an initial matter, OpenAI's request completely disregards the parties' agreed upon search protocol. In July 2024, the parties negotiated date ranges for collection and The Times represented that it would collect documents from its custodians beginning from January 1, 2022 (ChatGPT was publicly released in November 2022). Defendants did not object to this proposal. Subsequently, months later, OpenAI demanded that The Times go back and re-collect documents from five custodians for an expanded date range, and The Times did so. OpenAI's yet further demand that The Times once again conduct a collection of an additional four years for six of its most senior custodians' documents and Slack messages is unjustified.</p> <p>In any event, OpenAI's request to extend collection four years prior to the public release of ChatGPT makes no sense. Specifically, with respect to evidence of impact on The Times, <i>OpenAI</i> has documents relevant to its misconduct from well before 2022, when it used millions of The Times's works to develop its products. However, because much of this unlawful activity came before the public release of OpenAI's GPT product in 2022, The Times is unlikely to have custodial documents relevant to impact from this earlier period.</p> <p>Moreover, as The Times has repeatedly explained, it cannot run OpenAI's complex terms and obtain hit counts absent collection, processing, and ingestion, across emails (which can only accommodate simple search strings) and especially <i>Drive data</i> in Google Workspace (which cannot be searched in this manner at all). Similarly, The Times's ability to generate hit counts for Slack data without transferring the custodian's full files to a document vendor is limited, as searches time out where there is a significant volume of data. Indeed, despite OpenAI's suggestion to the contrary, The Times cannot use Google Vault to run</p>

Motion Dkt. No.	Date Filed	Opp. Dkt. No.	Topic in Dispute	Party Raising Dispute	Party Responding	Status of Dispute	Parties' Shorthand Positions
							<p>search terms across Slack data—as Slack messages are not housed in Google Workspace. Accordingly, OpenAI's assertion that all negotiated search "terms should be applied to [this] custodial collection" is a nonstarter.</p> <p>In an effort to resolve this issue, The Times ran a broad search across the relevant data for Google Mail messages relating to ChatGPT and OpenAI. None of the documents were responsive, including for OpenAI's statute of limitations defense or to the impact of its LLM products on The Times. Accordingly, the Court should deny OpenAI's request.</p>
250 - Sealed 251 - Public	4/1/25	253 (filed 4/3/25)	Additional CIR ESI Custodians	OpenAI	CIR	<p>OpenAI and CIR have submitted letter briefs, on which the Court has not yet heard argument.</p>	<p>OpenAI's position: OpenAI seeks an order compelling CIR to add as custodians Robert Wise, Marla Jones-Newman, Emily Harris, Ruth Murai, and CIR editors possessing documents relating to use of third-party works. CIR's three current custodians are insufficient to cover the over 80 RFPs for which it has agreed to produce responsive documents. In fact, CIR's CEO (and 30(b)(6) custodial deposition witness) admitted during her deposition that these proposed custodians possess "relevant and noncumulative" documents that are not otherwise in the possession of CIR's current custodians. <i>Coventry Cap. US LLC v. EEA Life Settlements, Inc.</i>, No. 17-cv-7417, 2021 WL 961750, at *2 (S.D.N.Y. Mar. 15, 2021); <i>see</i> Dkt. 250 at 2–3. These "relevant and noncumulative" documents relate to CIR's use of copyright management information, exclusion protocols, employment agreements, use of generative AI and third-party works, and financial analyses—all of which are central to the claims and defenses in this litigation. <i>Id.</i></p> <p>CIR cannot justify its unwillingness to add the proposed custodians. It has not articulated any specific burden it would allegedly suffer from the inclusion of additional custodians, and the size of its organization, alone, is not a sufficient basis for refusing to include custodians possessing "relevant and noncumulative evidence." <i>Coventry Cap.</i>, 2021 WL 961750, at *2. Likewise, the fact that CIR has named certain purported "decision-makers" as custodians does not excuse it from adding individuals for which CIR's CEO testified have responsive, noncumulative documents. Lastly, CIR's demand that OpenAI first review its <i>still ongoing</i> productions before seeking additional custodians amounts to nothing more than a delay tactic that CIR has employed throughout this litigation, including by providing the bulk of its document</p>

Motion Dkt. No.	Date Filed	Opp. Dkt. No.	Topic in Dispute	Party Raising Dispute	Party Responding	Status of Dispute	Parties' Shorthand Positions
							<p>production only two weeks ago and repeatedly failing to comply with the ESI order in responding to OpenAI's search term and custodian proposals. <i>See</i> Dkt. 251-1 at 1–2, 6; Dkt. 250-1 at 14. OpenAI has already been prejudiced by CIR's delays, and with the interim fact discovery deadline less than a month away, OpenAI cannot afford to wait any longer.</p> <p>Plaintiff's position: The 5 individuals that OpenAI moves for CIR to add as document custodians are not necessary, useful, or appropriate in this case. Indeed, CIR has already agreed to <i>manually</i> collect the majority of the discovery OpenAI seeks from these additional individuals (e.g., CIR's use of copyright management information, exclusion protocols, employment agreements, and use of generative AI and third-party works). To the extent CIR's manual collection does not sufficiently capture documents responsive to OpenAI's discovery requests, such additional materials are in the possession of CIR's existing custodians – its CEO, CFO, and VP of Media Sales/Business Development. CIR's 30(b)(6) custodial witness testified that these custodians are the key decisionmakers at CIR regarding any issues relevant to this litigation and are involved in all policy decisions that could have an impact on the strength of CIR's claims or OpenAI's defenses. OpenAI's allegations to the contrary do not support compelling additional custodians. It would be unduly burdensome, not proportional to the needs of the case, and (at best) duplicative to add the custodians identified in OpenAI's motion.</p>
III. Disputes Related to Plaintiffs' Requests for Inspection							
379 , Dkt. 483	1/13/25	420 , Dkt. 491	OpenAI's Log Output Data	News Plaintiffs	OpenAI	The Court heard argument at the January 22, 2025 conference, denied Plaintiffs' pre-motion letter at 379, and ordered the Parties to meet and confer. Dkt. 441.	<p>Plaintiffs' Position: In response to The Times's and Daily New Plaintiffs' renewed motion (Dkt. 483), OpenAI sent a letter to Plaintiffs on March 31, 2025: (1) admitting that OpenAI has destroyed an unacceptably large number of conversations (i.e., prompt and output data) for its ChatGPT Free, Pro, and Plus products since the December 27, 2023, filing of The Times's original complaint, and (2) refusing to suspend its data destruction policies or implement a solution to preserve only relevant output log data. OpenAI's March 31 letter, however, states that OpenAI does not track and thus does not have available the number of API Platform "completions" and ChatGPT Enterprise "conversations" that it has not retained. Plaintiffs believe that OpenAI's destruction of API completions and conversations from these two products is likely substantial in view of OpenAI's Enterprise Privacy policy, which states:</p>

Motion Dkt. No.	Date Filed	Opp. Dkt. No.	Topic in Dispute	Party Raising Dispute	Party Responding	Status of Dispute	Parties' Shorthand Positions
						Plaintiffs filed a separate pre-motion letter on March 25, 2025 at Dkt. 483.	<p>“OpenAI may securely retain API inputs and outputs for up to 30 days to provide the services and to identify abuse,” and “[a]fter 30 days, API inputs and outputs are removed from our systems, unless we are legally required to retain them.” <i>See</i> Enterprise privacy at OpenAI, dated October 31, 2024, available at https://openai.com/enterprise-privacy. Plaintiffs sent OpenAI a responsive letter on April 2, 2025, asking for details about the data OpenAI has retained for the API Platform and ChatGPT Enterprise products and for clarifications about the information OpenAI provided as to its ChatGPT Free, Pro, and Plus products, including why there were substantial spikes in deleted ChatGPT Free, Pro, and Plus conversations just before November 2024 and just after January 2025.</p> <p>Plaintiffs maintain their requests in ECF 379 and ECF 483 for the Court to order OpenAI to: (1) preserve its output log data, including output log data that contains News Plaintiffs’ content, hereinafter, and (2) identify all output log data that has been destroyed, when it was destroyed, and whether it is retrievable. In the meantime, Plaintiffs are continuing to confer with OpenAI on discovery into the output log data that it has preserved and are scheduling a call among the parties’ technical experts and witnesses to discuss sampling methodologies.</p> <p>OpenAI’s Position: There is no outstanding dispute for the Court to adjudicate with respect to Plaintiffs’ motion at ECF 483. Pursuant to the Court’s order in ECF 441, the parties continue to meet and confer “regarding the privacy and technological considerations implicated in Plaintiffs’ request” regarding user data retention.</p> <p>OpenAI has investigated and confirmed, based on its reasonable investigation, what data exists in connection with Plaintiffs’ demands and relayed that information to Plaintiffs in writing on March 31, 2025. OpenAI provided Plaintiffs with context for the voluminous number of conversations that have been retained, and informed Plaintiffs of the estimated fraction of ChatGPT Free, Pro, and Plus conversations that have not. OpenAI has also confirmed for Plaintiffs that the ChatGPT Free, Pro, and Plus conversation data that has not been retained is not retrievable, and that it does not track such data for the API Platform and ChatGPT Enterprise.</p>

Motion Dkt. No.	Date Filed	Opp. Dkt. No.	Topic in Dispute	Party Raising Dispute	Party Responding	Status of Dispute	Parties' Shorthand Positions
							OpenAI looks forward to continuing to work with Plaintiffs to find a path forward that ensures user privacy interests are protected, while permitting Plaintiffs access to the vast quantities of ChatGPT conversations that are available for sampling and analysis.
Dkt. 487	March 31, 2025	508	News Plaintiffs seek discovery into all versions of Microsoft's "Copilot" product and inspection of the "Deucalion" LLM	News Plaintiffs	Microsoft	The dispute is fully briefed and the Court has not yet heard argument	<p>Plaintiffs' Position: This motion to compel addresses two issues relating to Microsoft.</p> <p><i>First</i>, Plaintiffs request discovery into all versions of Microsoft's "Copilot" product. Plaintiffs recently learned that Microsoft has not been providing discovery into its current version of "Copilot"—which was launched October 1, 2024—on the ground that it is a new product that is not covered by the allegations in the Complaints. But the current version of Copilot has the same name as the Copilot product identified in the Complaints, uses the same OpenAI LLMs identified in the Complaints, and performs the same functionality as the Copilot referenced in the Complaints, including regurgitating Plaintiffs' copyrighted articles. Microsoft's suggestion that Plaintiffs have known about what it calls "New Copilot" since an October 1, 2024, blog post is also mistaken. The blog post does not say that the original Copilot was deprecated, identify changes in the underlying system architecture, or characterize the October 1, 2024, Copilot as fundamentally different product. Instead, the announcement describes a "refreshed" Copilot. Accordingly, the current version of Copilot is not a "new" product that was not identified in the Complaints. Accordingly, Plaintiffs request that the Court grant Plaintiffs' requested discovery into all versions of Microsoft's "Copilot" product, including: (1) updating its responses to News Plaintiffs' interrogatories to encompass information about the refreshed Copilot; (2) searching for and producing documents responsive to News Plaintiffs' requests for production that relate to the refreshed Copilot, including producing documents that were withheld on the basis that they relate to the refreshed Copilot; and (3) producing output log data, source code, training data, and any models for the refreshed Copilot in response to News Plaintiffs' requests for inspection.</p> <p><i>Second</i>, Plaintiffs request access to inspect Microsoft's "Deucalion" LLM, which is based off an OpenAI GPT model that Microsoft fine-tuned at OpenAI and commercially deployed in the Copilot product deprecated on October 1, 2024. Plaintiffs have already conducted such</p>

Motion Dkt. No.	Date Filed	Opp. Dkt. No.	Topic in Dispute	Party Raising Dispute	Party Responding	Status of Dispute	Parties' Shorthand Positions
							<p>an inspection with respect to OpenAI's commercially available LLMs, but Microsoft asserts that Plaintiffs should not be entitled access to inspect Deucalion because the model's outputs were mediated through other layers of software. But irrespective of end user's direct access to the model, the model itself is relevant to News Plaintiffs allegations concerning model memorization, including: (1) the model itself is infringing and not transformative (<i>e.g.</i>, because it encodes copies of the asserted works), and (2) memorized (or regurgitated) outputs are further evidence of the extent and nature of Defendants' unauthorized use of News Plaintiffs' content during training and as part of Copilot's outputs. Microsoft has also not articulated any undue burden in making Deucalion available for inspection. Accordingly, the Court should grant Plaintiffs' requested inspection of Microsoft's "Deucalion" LLM.</p> <p>Microsoft's Position: News Plaintiffs' motion concerning two unrelated discovery requests should be denied; indeed, the first was already denied.</p> <p><i>First</i>, Microsoft's Copilot-branded product that was released on October 1, 2024 ("New Copilot") is a new product that did not exist at the time of the Complaint, was built on an entirely different architecture, involves a separate team, and would necessitate starting over in discovery. News Plaintiffs have known about New Copilot since it was launched, included a blog post about it in The Times's request in October 2024 to expand the scope of discovery, and the Court denied it. <i>See</i> ECF 365 at 5. As the Court recognized during the December 2024 hearing, if News Plaintiffs want discovery on New Copilot, they can either file a new lawsuit or seek leave to amend under Rules 15 and 16. Adding New Copilot to this case would dramatically increase the scope of discovery, entailing a new set of source code, business documents, technology documents, and ESI from an entirely different set of custodians. It would be like starting over, making an already expansive case unmanageable. News Plaintiffs' renewed request for discovery into New Copilot should be denied.</p> <p><i>Second</i>, News Plaintiffs' demand to test a bare model called Deucalion is disproportionate to the needs of the case and should be denied. Deucalion is a fine-tuned version of GPT3.5 Turbo that was never made available to the public in the form in which News Plaintiffs seek to test</p>

Motion Dkt. No.	Date Filed	Opp. Dkt. No.	Topic in Dispute	Party Raising Dispute	Party Responding	Status of Dispute	Parties' Shorthand Positions
							it. There were and are myriad other options for News Plaintiffs to get at relevant information about the model, including the fact that they had freely available access to Copilot while the Deucalion model was in operation throughout most of 2024, they have already inspected the training data used to train the GPT3.5Turbo model on which it was based, and they also have the ability to sample user output data to understand the actual uses by consumers of the Deucalion model in operation. News Plaintiffs equate their request to inspect Deucalion to the inspection of OpenAI's models, but this is misleading. The models OpenAI provided for inspection were publicly available in the form inspected, which is not what is sought here. Rather, Plaintiffs seek to force Microsoft to provide it without the metaprompts, filters, and output mitigations that were all part of the safety features of Copilot. Testing of the bare model does not correspond to its operation in Copilot. Worse, providing the bare model would enable News Plaintiffs to radically distort the model's behavior through adversarial prompting, causing significant prejudice. Adversarial prompting of models has already been an issue in these proceedings, and the issue should not be compounded further. There were and are other avenues to get at any potentially relevant information. News Plaintiffs' request to inspect a model in a form it was never offered to the public should be denied.
305 ¹ 328	11/1/24 11/20/24	305 345	OpenAI's Training Data	News Plaintiffs	OpenAI	The Court heard argument on December 3, 2024, and on January 22, 2025.	Plaintiffs' position: OpenAI ran Plaintiffs' "URL Search" of OpenAI training data and produced the search results to The Times and Daily News Plaintiffs on January 29, 2025. Plaintiffs are continuing to investigate these results. As directed by the Court, the parties met and conferred regarding proposed datasets and threshold values for the proposed "N-Gram Search" of OpenAI training data and have reached agreement. OpenAI has now sent portions of the code for the "N-Gram Search" they expect to run to Plaintiffs, and Plaintiffs have raised an issue for OpenAI to address regarding how the code works. Specifically, OpenAI's code in effect doubles the n-gram threshold that the parties agreed to—requiring that a record have copied 106 words (100 6-grams) before a match could occur, rather than the 50 n-gram threshold agreed to. Plaintiffs expect OpenAI to address the issue they have raised and promptly run the search as previously agreed and produce results by April 24, but OpenAI has not

¹ Unless otherwise indicated, all docket numbers are from *New York Times Company v. Microsoft Corp., et al.*, Case No. 1:23-cv-11195-SHS.

Motion Dkt. No.	Date Filed	Opp. Dkt. No.	Topic in Dispute	Party Raising Dispute	Party Responding	Status of Dispute	Parties' Shorthand Positions
							<p>provided any information about when the results will be provided or committed to a date certain. Given the extensive amount of time for the Parties to reach this point, Plaintiffs request that OpenAI produce results by April 24. OpenAI frames this as a “last-minute request,” but almost immediately after receiving OpenAI’s code for the n-gram search, Plaintiffs made this request in order to preserve the parties’ prior agreement. OpenAI has not identified any reason why it could not address the issue that Plaintiffs have raised, and Plaintiffs’ understanding is that there are no technical reasons preventing it.</p> <p>OpenAI’s Position: At the last status conference, the parties represented that all but two final N-Gram Search details had been resolved. Two weeks later, they reached agreement on those issues as well. ECF 462. Because the parties were in agreement that (1) OpenAI would implement Plaintiffs’ search code, but (2) Plaintiffs’ code needed to be rewritten to work at scale on OpenAI’s systems, OpenAI’s engineers have been working to refactor Plaintiffs’ original code. As a final pre-flight check, OpenAI shared the refactored code to make sure that its engineers had implemented Plaintiffs’ method correctly.</p> <p>The “issue” that Plaintiffs have now raised is that they want OpenAI to consider changing one of the key parameters of the search. Plaintiffs originally included that parameter in a detailed set of instructions and code for “how the n-gram search should take place.” In other words, OpenAI implemented this parameter <i>exactly</i> as Plaintiffs had always requested, which was what the parties had agreed to. Plaintiffs write that they are trying to “preserve the parties’ prior agreement,” but in actuality, they are trying to renegotiate the terms that they explicitly told OpenAI to adopt. OpenAI is willing to consider Plaintiffs’ last-minute request, but this parameter (which Plaintiffs have always asked OpenAI to use) was integral to OpenAI’s agreement on other key details. Rather than re-open the parties’ agreement on those details as well, OpenAI is prepared to proceed with the search that was agreed upon.</p>
IV. Disputes Related to OpenAI’s Document Productions							
Dkt. 489	March 31, 2025	507	The Times seeks to compel OpenAI to produce discrete financial	NYT	OpenAI	The Times filed a similar motion in advance of the January 22,	Plaintiffs’ Position: The Times on March 31 renewed its January 13, 2025 request for an order requiring OpenAI to produce discrete financial documents and investor materials. The Times has substantially narrowed its request, consistent with this Court’s instruction “to think about whether there are particular investments or transactions or particular

Motion Dkt. No.	Date Filed	Opp. Dkt. No.	Topic in Dispute	Party Raising Dispute	Party Responding	Status of Dispute	Parties' Shorthand Positions
			documents and investor materials concerning OpenAI's valuation and anticipated profits			2025 conference (Dkt. 398). Following argument, the Court ordered the parties to continue to meet and confer. Dkt. 441. The Times has now renewed its motion.	<p>presentations or other documents . . . that you might want to ask OpenAI to focus on" and to "try to address OpenAI's concern about the request being overbroad." Jan. 22, 2025 Hearing Tr. at 87:12-22. The Court should now adopt The Times's compromise proposal, for OpenAI to produce: (1) documents reflecting internal evaluations of its anticipated profits and OpenAI's valuation, and (2) documents provided to the following OpenAI investors that discuss OpenAI's anticipated profits and valuation: Microsoft, MGX, Thrive Capital, Softbank, and Tiger Global.</p> <p>This compromise proposal targets information The Times has not otherwise received. This discovery is also highly relevant, for at least four issues: (i) the first fair use factor—the purpose and character of the use, (ii) OpenAI's ability to pay for copyrighted content, (iii) willfulness and motive, and (iv) damages. For example, as for fair use factor one, "[t]he greater the private economic rewards reaped by the secondary user (to the exclusion of broader public benefits), the more likely the first factor will favor the copyright holder and the less likely the use will be considered fair." <i>Am. Geophysical Union v. Texaco Inc.</i>, 60 F.3d 913, 922 (2d Cir. 1994). Yet OpenAI refuses to provide basic discovery into "the private economic rewards" that it anticipates "reap[ing]" for itself and its employees, thus preventing The Times (and the Court) from balancing those "private economic rewards" against the alleged "public benefits" that OpenAI touts. <i>Texaco</i>, 60 F.3d at 922. OpenAI below incorrectly suggests The Times has not cited any cases ordering such discovery in a copyright case. <i>E.g.</i>, <i>Daily Palm, LLC v. Bird Rides, Inc.</i>, 2021 WL 6496853, at *1 (C.D. Cal. Oct. 18, 2021) (crediting plaintiffs' argument that disgorgement can "be measured by any increase in the valuation of Defendant caused by Defendant's alleged infringement of Plaintiff's copyrighted work," and ordering discovery into the "valuation of defendant").</p> <p>Finally, the burden to OpenAI is minimal. For category (1) OpenAI has admitted these documents are "discrete" and has therefore not even raised any burden argument. And for category (2), OpenAI has already located the "data rooms" it used to share documents with most of these investors, meaning that OpenAI already knows exactly where to find the responsive materials. Tellingly, OpenAI below raises no burden argument, which</p>

Motion Dkt. No.	Date Filed	Opp. Dkt. No.	Topic in Dispute	Party Raising Dispute	Party Responding	Status of Dispute	Parties' Shorthand Positions
							<p>underscores that the parties have a ripe dispute about the relevance of the requested materials.</p> <p>OpenAI's Position: The Times seeks to compel OpenAI to produce documents that courts have consistently held are not discoverable in copyright actions.</p> <p>Courts have uniformly refused to compel production of a copyright defendant's financial projections as irrelevant and disproportionate. <i>See In re ChatGPT</i>, Case No. 23-cv-03223-AMO, Dkt. 247 at 2 (N.D. Cal. Jan. 13, 2025); <i>Phoenix Techs., Ltd. v. VMware, Inc.</i>, 2016 WL 5725044, at *2 (N.D. Cal. Sept. 30, 2016). They have also held that a company's valuations are not discoverable for similar reasons. <i>Bluewater Music Servs. Corp. v. Spotify USA Inc.</i>, 2018 WL 10152456, at *7 (M.D. Tenn. July 10, 2018). And they have rejected arguments that business documents (like the presentations The Times demands here) are relevant simply because they discuss the defendant's financial information. <i>See In re ChatGPT</i>, Dkt. 247 at 2; <i>Kaseberg v. Conaco, LLC</i>, 2016 WL 3997600, at *7–8 (S.D. Cal. July 26, 2016).</p> <p>Indeed, The Times cannot identify a single case compelling the scope of production they presently demand. The only case they cite concerning discovery in a copyright action (<i>Daily Palm, LLC v. Bird Rides, Inc.</i>) addressed the purely legal question of what damages a copyright plaintiff could seek as a matter of law—not the arguments OpenAI has raised here. And, even then, the court ordered far more limited discovery than The Times requests.</p> <p>OpenAI has repeatedly offered to meet and confer about conducting additional reasonable searches for documents relating to relevant, substantive issues that aren't addressed by OpenAI's existing offers to produce, but The Times has flatly ignored those offers and refused to identify any information it is seeking from the requested materials beyond OpenAI's projections and valuations.</p> <p>OpenAI's projections and valuations are not relevant, and The Times's assertions to the contrary are flimsy, at best. Documents regarding OpenAI's overall financial projections and valuation are not relevant to</p>

Motion Dkt. No.	Date Filed	Opp. Dkt. No.	Topic in Dispute	Party Raising Dispute	Party Responding	Status of Dispute	Parties' Shorthand Positions
							<p>fair use factor one because they don't concern the allegedly infringing use. "The question under [fair use] factor one is the purpose and character of the use, not of the alleged infringer." <i>Byrne v. British Broad. Corp.</i>, 132 F. Supp. 2d 229, 234 (S.D.N.Y. 2001) (Stein, J.).</p> <p>Regarding "ability to pay," OpenAI has already produced documents that would reflect its "ability to pay for copyrighted content." And, in any event, The Times cites no authority for the proposition that projections and valuations are relevant to inquiry or that such an inquiry is relevant to their copyright claims.</p> <p>The Times's assertion that the requested documents are relevant to OpenAI's supposed motive to willfully infringe contravenes the law on willfulness, which establishes that "one's motive to infringe does not bear on the willful nature of his infringement." <i>Compaq Comput. Corp.</i>, 2001 WL 34104827, at *1 (citing <i>Fitzgerald Publ'g Co., Inc. v. Baylor Publ'g Co., Inc.</i>, 807 F.2d 1110, 1115 (2nd Cir. 1986)).</p> <p>And, contrary to what The Times asserts, courts have consistently held that financial projections and valuations are not relevant to damages. <i>In re ChatGPT</i>, Dkt. 247 at 2; <i>Phoenix Techs. Ltd.</i>, 2016 WL 5725044, at *2; <i>Bluewater Music Servs. Corp.</i>, 2018 WL 10152456, at *7. Even if that information might be relevant, The Times cannot explain how they would use OpenAI's financial projections/valuation to calculate damages arising from the allegedly infringing use of their copyrighted materials. In the absence of such an explanation, the requested discovery is improper. This, too, courts have consistently held. <i>See, e.g., Structured Asset Sales, LLC v. Sheeran</i>, 433 F. Supp. 3d 608, 612 (S.D.N.Y. 2020).</p> <p>The Times' assertion that OpenAI lacks a burden concern is perplexing. OpenAI has cited several cases holding that searching for and producing the documents The Times requests is not proportionate to the needs of the case and urged this Court to follow their guidance. The Times offers no serious argument rebutting the holdings of those numerous cases.</p>
322 -Sealed	11/19/24	339	The Times's RFP 2 (documents OpenAI has	NYT	OpenAI	The Court heard argument on this motion at	Plaintiff's position: The Times maintains its request for an order requiring OpenAI to identify any relevant domestic proceedings in which it has produced documents related to the issues in this case.

Motion Dkt. No.	Date Filed	Opp. Dkt. No.	Topic in Dispute	Party Raising Dispute	Party Responding	Status of Dispute	Parties' Shorthand Positions
323 -Public			submitted to domestic governmental entities relating to the allegations in the Complaint)			the January 22, 2025 conference, and ordered the parties to meet-and-confer and to provide a status update in the February 13 chart (Dkt. 462)	<p>The Court should now adopt The Times's compromise proposal, as reflected in the February 13 chart. Dkt. 462. As a first step, OpenAI should produce any interrogatory responses or responses to written questions that OpenAI submitted for the Delaware Attorney General's ongoing investigation into OpenAI's transition to a for-profit company. https://www.reuters.com/technology/artificial-intelligence/openais-altman-says-musk-offer-is-tactic-to-try-mess-with-us-2025-02-11/. The Times is targeting this investigation because it addresses "OpenAI's proposed changes to ensure the company is 'adhering to its specific charitable purposes for the benefit of the public beneficiaries, as opposed to the commercial or private interests of OpenAI's directors or partners.'" <i>Id.</i> The requested documents are therefore relevant to fair use factor one, particularly insofar as they reveal the extent to which OpenAI is prioritizing commercialization. Indeed, commercialization is not a binary proposition. "The greater the private economic rewards reaped by the secondary user (to the exclusion of broader public benefits), the more likely the first factor will favor the copyright holder and the less likely the use will be considered fair." <i>Am. Geophysical Union v. Texaco Inc.</i>, 60 F.3d 913, 922 (2d Cir. 1994). This discovery will shed light on whether OpenAI is prioritizing "private economic rewards" over "broader public benefits." <i>Id.</i> Tellingly, OpenAI below does not claim it would be burdensome to comply with The Times's request, which again focuses on just one highly relevant proceeding and on information The Times has not otherwise received.</p> <p>OpenAI's Position: The Court previously rejected The Times's motion to compel with regard to documents submitted to foreign regulators (ECF 398), and it should deny this parallel motion as well.</p> <p>The Times has issued 155 RFPs in this matter, which is 25 more RFPs than had been issued when the parties last submitted a chart on this dispute. See Dkt. 462. The pending RFPs canvass all conceivably relevant topics. As relevant here, The Times has also requested all documents submitted to domestic governmental entities, without regard to whether those materials have been produced in response to other RFPs or relate to substantive meet-and-confer efforts. The Court previously encouraged The Times to focus its discovery requests on "underlying information . . . relevant to the case," not on materials that might have been deemed relevant in some separate context. (Transcript, Jan. 22, 2025, at 75:1-5). Indeed, with regard to alleged "commercial nature" of</p>

Motion Dkt. No.	Date Filed	Opp. Dkt. No.	Topic in Dispute	Party Raising Dispute	Party Responding	Status of Dispute	Parties' Shorthand Positions
							<p>OpenAI's business, The Times has already issued an RFP focused on precisely that issue. See RFP 11. That RFP has been the subject of motion practice and extensive meet-and-confer discussions. The Times should not be permitted to end-run around those negotiations simply by pointing to public reporting of a state regulator's entirely distinct inquiry. This is particularly true when the alleged relevance is a corporate change that has not even occurred yet, and where The Times blatantly misstates the alleged relevance - there is no dispute that OpenAI has had a for-profit entity since 2019. The Times is engaged in a fishing expedition, and its motion to compel should be denied.</p> <p>The Times's suggestion that it is offering a narrowed resolution to this RFP is false. It has clarified in meet-and-confer discussions that it expects OpenAI to canvas other, unidentified regulatory engagements for potentially relevant materials. Instead of casting about for materials potentially deemed relevant in other contexts, The Times should, as this Court directed, focus on "underlying information . . . relevant to the case." (Transcript, Jan. 22, 2025, at 75:1-5).</p>
V. Updates for Previously Briefed & Argued Disputes Between News Plaintiffs and Microsoft							
311	11/18/24	338	Microsoft's Log Output Data	News Plaintiffs	Microsoft	<p>The Court heard argument at the January 22, 2025 conference, and ordered the Parties to meet and confer and to provide a status update in the February 13 chart (Dkt. 462).</p>	<p>Plaintiffs' position: On February 10, 2025, technical experts for The Times and Daily News Plaintiffs attended an informal discussion with Microsoft's outside consultants. No Microsoft employees attended the discussion, and because Microsoft's outside consultants' knowledge of Microsoft's log data was limited to the same materials produced to Plaintiffs, the outside consultants could not answer many of the questions raised during the discussion and previously outlined Plaintiffs' letter correspondence. On February 12, The Times and Daily News Plaintiffs sent a letter to Microsoft to follow up on items that the consultants were not able to answer during the informal discussion and to propose a sampling methodology, and requested responses by February 17. Microsoft has yet to respond to that letter. News Plaintiffs will continue to meet and confer with Microsoft on discovery into the log data but request that the Court order a date certain by which: (1) Microsoft responds to Plaintiffs' February 12 letter, (2) the Parties complete their meet and confer as to sampling methodologies of the log data, and (3) production of Microsoft produces the log data according to the agreed upon (or Court ordered) sampling methodologies.</p>

Motion Dkt. No.	Date Filed	Opp. Dkt. No.	Topic in Dispute	Party Raising Dispute	Party Responding	Status of Dispute	Parties' Shorthand Positions
							<p>Microsoft's Position: Plaintiffs' initial motion was denied in ECF 365. Since that time, the parties have been working together in good faith to address these issues. The parties held an initial extensive meet and confer session. Microsoft disagrees that its experts were unprepared for that session. Rather, Plaintiffs spent nearly half of the time asking questions that were not in the scope of topics sent in advance of the meeting. Moreover, when follow-up questions related to the topics were asked, the experts were able to answer most questions. Microsoft has been working diligently to answer the additional questions raised in Plaintiffs' February 12 letter. Counsel anticipates being able to provide a written response in the next week. At the same time, Microsoft has also been working on streamlining the process by creating faster and more efficient means of accessing the data, both in order to answer Plaintiffs' questions and in support of the ultimate project of sampling the data. Because Microsoft has been working on multiple fronts to create a better process for understanding and sampling the data, it has taken more time than anticipated to answer Plaintiffs' questions.</p> <p>After Plaintiffs review the responses to their February 12 letter and after the MDL is organized so that all data matters can be addressed on a uniform basis, the parties should have another informal conversation about log data. The log data is stored in an unstructured format and therefore reporting scripts must be run on data to create samples. This process has proven quite challenging as has the overall effort to understand the data. A further meet and confer session will be productive in enabling Plaintiffs to formulate their sampling strategy. After that, sampling efforts are likely to require testing and iteration before any approach can be confirmed. There is no need for deadlines regarding the ongoing process of addressing log data to be set at this time.</p>
396	1/13/25	407	The Times's RFPs 61 (documents concerning licensing negotiations)	NYT	Microsoft	The Court heard argument on this motion at the January 22, 2025 conference, and ordered the parties to	<p>Plaintiff's position: The Times has substantially narrowed its request for licensing negotiation documents for deals that were never executed, which The Times has been seeking since January. Because Microsoft has still not committed to producing any such documents, The Times respectfully asks the Court to adopt its compromise proposal, summarized below.</p> <p>These requested documents and communications are relevant to The Times's allegation that there is a market for the at-issue data. FAC ¶ 155;</p>

Motion Dkt. No.	Date Filed	Opp. Dkt. No.	Topic in Dispute	Party Raising Dispute	Party Responding	Status of Dispute	Parties' Shorthand Positions
						meet-and-confer and to provide a status update in the February 13 chart (Dkt. 462)	<p><i>see A&M Recs., Inc. v. Napster, Inc.</i>, 239 F.3d 1004, 1016 (9th Cir. 2001) (“Fair use, when properly applied, is limited to copying by others which does not materially impair the marketability of the work which is copied.” (quoting <i>Harper & Row Publishers, Inc. v. Nation Enters.</i>, 471 U.S. 539, 566-67 (1985)). These communications can be just as probative of the existence of a market for content as negotiations that did result in an agreement. For example, if communications show that Microsoft was willing to pay a news organization a certain amount of money to use their content for Microsoft Copilot, that admission would be relevant to assessing the market for news data, even if that deal never materialized.</p> <p>At the January 22 hearing, this Court instructed The Times to “narrow the scope of [the request] to what the market might be in the News cases.” Jan. 22, 2025 Tr. at 103:7-9. Consistent with that guidance, The Times now seeks documents and communications regarding licensing negotiations that did not result in an executed agreement, but limited to (1) Journalism content providers, as the parties have agreed to define that term to resolve other RFPs, and (2) negotiations that moved far enough along where a party prepared a term sheet. The Times has also agreed to narrow the scope of the models and products for this RFP (A) the GPT-based models indisputably at issue in the case, and (B) the following Microsoft generative AI products: Bing Chat, Microsoft Copilot (including all versions), and Microsoft 363 Copilot.</p> <p>Microsoft’s Position: Plaintiffs’ initial motion was denied in ECF 441, but the parties have continued to meet and confer. As an initial matter, The Times’s apparent introduction of “all versions” of Copilot, which Microsoft takes to refer to the newer product that is the subject of the briefs at ECF 487 and 508, is a step backward, as it expands the scope of the request when the parties are trying to narrow and get to agreement. Indeed, Microsoft asked in the parties’ most recent written correspondence on this issue, on March 31, to confirm that The Times’s current proposal was “still limited to the models and products at issue in the case (setting aside, ... our dispute about what products are subject to discovery, as set out in your motion filed today).” The Times did not respond; it appears the answer is no.</p> <p>As Microsoft has communicated to The Times, Microsoft will be producing additional documents in response to this RFP beyond what is set out in Microsoft’s served written response. Microsoft is continuing to</p>

Motion Dkt. No.	Date Filed	Opp. Dkt. No.	Topic in Dispute	Party Raising Dispute	Party Responding	Status of Dispute	Parties' Shorthand Positions
							evaluate the scope of those additional documents and whether such documents would be responsive to the scope articulated in The Times's previous proposal. Since The Times has just now expanded its request to include "all versions" of Copilot, Microsoft has not been able to evaluate that request. For all these reasons, the parties should continue to meet and confer, and no relief is warranted at this time.